Urban Overtourism Detection Based on Graph Temporal Convolutional Networks

Xiangjie Kong, *Senior Member, IEEE*, Zhiqiang Huang, Guojiang Shen, Hang Lin, and Mingjie Lv

***Abstract*— Urban overtourism results in heavy traffic, degraded tourist experiences, and overloaded infrastructure. Detecting urban overtourism at the early stage is important to minimize the adverse effects. However, urban overtourism detection (UOD) is a challenging task due to ambiguity, sparsity, and complex spatiotemporal relations of overtourism. In this article, we propose a novel UOD framework based on graph temporal convolutional networks (TCNs) to tackle the challenges mentioned above. More specifically, we propose the grid overtourism mode (GOM) to detect urban overtourism on a grid level and propose the overtourism detection mechanism, which gives a quantitative definition of overtourism and screens out the regions where overtourism may occur as candidate regions. Then, we construct the GOM graphs of the candidate regions. Next, we employ the graph TCNs to model the complex spatiotemporal relations of urban overtourism and predict the future GOM graph at the next time interval. Finally, we calculate the urban overtourism scores based on the prediction results. The experiments are conducted based on a real-world dataset. The evaluation results demonstrate the effectiveness of our methods.**

***Index Terms*— Anomaly detection, graph temporal convolutional networks (TCNs), trajectory data mining, urban overtourism.**

I. Introduction

THE potential benefits of tourism are obvious. However, after decades of virtually uncontrolled growth, tourism now demonstrably creates more problems than benefits. The negative environmental and sociocultural impacts of tourism are becoming more evident and have resulted in the development of the term "overtourism."

The United Nations World Tourism Organization has defined overtourism as "the impact of tourism on a destination, or parts thereof, that excessively influences the perceived quality of life of citizens and/or quality of visitors' experiences in a negative way" [1]. Overtourism refers simply to the notion that there are too many visitors in a particular region. The negative impacts of overtourism include alienated residents, degraded tourist experiences, overloaded infrastructure, damage to nature, and threats to culture and heritage. In addition, with COVID-19 sweeping the globe, tourists need to maintain social distance while traveling, which gives new meaning to overtourism [2]. These impacts have led to increased concern from residents, with protests and reactions observed in major tourism destinations, such as Venice and Barcelona.

In recent years, how to mitigate the negative effects of overtourism has attracted extensive research interest from scholars. Katsumi et al. [3] proposed a method to discover a generic point of interest (POI), which is an alternative sightseeing spot potentially attractive enough for tourists to replace a well-known sightseeing spot. However, tourists will choose those truly famous tourist attractions that are easily accessible and functional, rather than those that are not easily accessible, underdeveloped, and just look like famous tourist attractions. To avoid the concentration of tourists in overtourism areas, it is possible to inform tourists of the congestion information in the destination in advance. Murata and Totsuka [4] developed an agent-based simulation model to estimate the effectiveness of "congestion information at the destination" among tourists. However, they simply consider overtourism according to traffic congestion. Traffic congestion is only the result of overtourism, and overtourism cannot be inferred from traffic congestion because there are many other reasons for traffic congestion such as traffic accidents and other emergencies.

The essence of overtourism is "the tragedy of the commons." The contradiction between infinite personal desires and limited tourism resources can lead to the destruction and depletion of tourism resources, as shown in Fig. 1. Therefore, it is of great value for city administrators and tourist attractions to detect urban overtourism in time and even predict it before happening. If most urban overtourism can be detected or predicted at the early stage, tourist attractions will have more time to formulate the corresponding measures and prevent serious incidents from occurring, such as limiting the scale of tourist reception and evacuating tourists in advance. In this way, it will not only save tourists a lot of time on the commute but also further improve their travel quality [5].

With the construction of the smart city and the rapid development of the intelligent transportation system, we have access to a large amount of heterogeneous multisource urban data, such as vehicle trajectories and geographic information system (GIS) data [6], [7]. The potential operation patterns of cities can be extracted through big data and artificial

Manuscript received 16 June 2022; revised 19 September 2022 and 2 November 2022; accepted 29 November 2022. This work was supported in part by the Zhejiang Provincial Natural Science Foundation under Grant LR21F020003, in part by the National Natural Science Foundation of China under Grant 62072409 and Grant 62073295, in part by the "Pioneer" and "Leading Goose" Research and Development Program of Zhejiang under Grant 2022C01050, and in part by the Fundamental Research Funds for the Provincial Universities of Zhejiang under Grant RF-B2020001. *(Corresponding author: Guojiang Shen.)*

Xiangjie Kong, Zhiqiang Huang, Guojiang Shen, and Hang Lin are with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: xjkong@ieee.org; zqhuang23@outlook.com; gjshen1975@zjut.edu.cn; hanglin_1999@outlook.com).

Mingjie Lv is with the Research Center for Intelligent Society and Governance, Zhejiang Lab, Hangzhou 311100, China (e-mail: lmj@zhejianglab.com).

Digital Object Identifier 10.1109/TCSS.2022.3226177

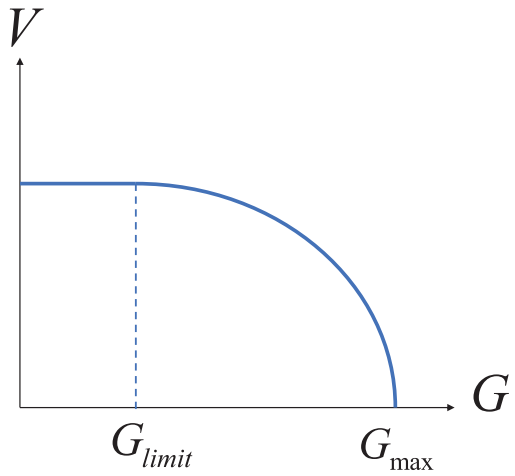


Fig. 1. Relation between the number of crowds and the per capita resources. When the number of crowds G exceeds the carrying capacity G_{limit} of the region, the per capita resources V will be reduced. When the number of crowds G reaches its maximum $G_{\max}$, there will be irreversible damage to the region.

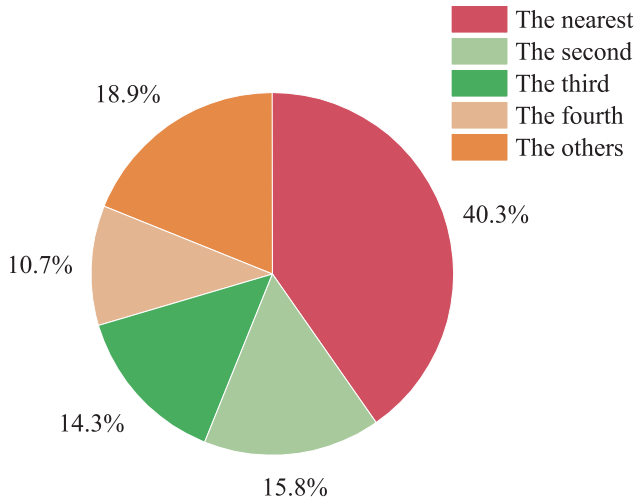


Fig. 2. Next destination for tourists starting from a grid in a day. We screen out all ODs that boarded from a grid and whose drop-off point was also a possible overtourism grid. Most tourists choose the nearest or more recent tourist attraction as their next destination.

intelligence algorithms from these data, which provides data support for many real-world applications such as urban anomaly detection. The purpose of urban anomaly detection is to find abnormal events that will affect the normal operation of the city, including abnormal traffic and crowd flow. However, overtourism is a special urban anomaly, and there are the following critical challenges for urban overtourism detection (UOD).

1) *Ambiguity of Urban Overtourism:* Urban overtourism is hard to be directly observed and understood timely. The criteria of urban overtourism may vary with region. Currently, there is a lack of quantitative indicators for overtourism.
2) *Sparsity of Urban Overtourism:* Compared with the whole urban regions, the regions where urban overtourism occurs only account for a small part. The extreme sparsity of labeled data makes it difficult to identify the characteristics of urban overtourism.
3) *Complex Spatiotemporal Relations of Urban Overtourism:* In the spatial dimension, after visiting a tourist attraction, tourists often choose a tourist attraction with a closer distance as their next destination, as shown in Fig. 2. In the temporal dimension, overtourism has a clear relationship with previous moments. All the spatiotemporal relations need to be considered when detecting urban overtourism.

To tackle the above challenges, we propose a novel UOD framework based on graph temporal convolutional networks (TCNs). To overcome the ambiguity of urban overtourism, we propose the grid overtourism mode (GOM) to detect urban overtourism on a grid level, which combines the traffic congestion index and the number of crowds because we believe that overtourism can be considered only when the increase in crowds leads to traffic congestion. We propose the overtourism detection mechanism to give a quantitative definition of overtourism. To overcome the sparsity of urban overtourism, we first divide the city into grids. Then, we screen all the grids twice to find the grids where overtourism may occur. We first screen the grid containing tourist attractions according to the latitude and longitude of tourist attractions in POI data. Then, we screen grids that have at least one overtourism according to the overtourism detection mechanism as candidate grids. We construct the GOM graph based on the screened girds, which allows us to focus on the grids where overtourism may occur. To study the complex spatiotemporal relations of urban overtourism, we employ graph TCNs to extract essential spatiotemporal relations. The main contributions of this article are given as follows.

1) To the best of our knowledge, we are the first to formalize the UOD problem. We propose the GOM to detect urban overtourism on a grid level and propose the overtourism detection mechanism to give a quantitative definition of overtourism.
2) We propose a novel UOD framework to screen out grids where overtourism may occur to construct GOM graphs and model the spatiotemporal characteristics of overtourism to achieve overtourism detection.
3) We conducted reliable experiments with a real-world dataset containing multisource data to demonstrate the effectiveness of the proposed methods.

We will present our work in the following order. First, we present in Section II the work related to our study that has inspired us. Then, to describe our framework more clearly, we introduce some preliminary concepts and also describe the framework of UOD in Section III. Furthermore, we introduce our framework in detail in Section IV. We present the details of experiments with a real-world dataset and verify the effectiveness of the proposed methods in Section V. We describe possible future research directions for this work in Section VI. Finally, we summarize this work in Section VII.

II. Related Work

In this section, we list two main aspects of the research background relevant to our study. Overtourism is a special urban anomaly, and we can study urban overtourism by mining

trajectory data. Thus, we will review the related work about trajectory data mining and urban anomaly detection.

A. Trajectory Data Mining

Trajectory data provide new data support and research ideas for studying the spatial movement of groups from the perspective of individuals. It is an important way to solve urban problems to explore the movement rules and activity patterns of human beings through trajectory data and then explore the deep knowledge contained therein. As the number of trajectory data increases rapidly, traditional data mining methods, especially statistics-based methods, could perform poorly when applied to trajectory data mining. Recently, deep learning models have achieved remarkable success in many domains due to the powerful ability of automatic feature representation learning and are also widely applied in various trajectory data mining tasks. Zheng [8] conducted a systematic survey of the major research into trajectory data mining, providing a panorama of the field as well as the scope of its research topics. Trajectory data are essentially spatiotemporal data, with both temporal and spatial attributes. Wang et al. [9] summarized the characteristics of spatiotemporal data, common processing methods, and related applications. To learn human mobility knowledge from fixed travel behaviors, Kong et al. [10] proposed a multipattern passenger flow prediction framework based on graph convolutional network. Cesario et al. [11] presented an integrated algorithm that supports the overall trajectory pattern discovery process for detecting user's mobility behaviors. They present a next-place prediction framework, which exploits location-based social networks (LBSNs) data to forecast the next location of an individual based on the observations of her mobility behavior over some period of time and the recent locations that she has visited and on global mobility in the considered geographic area [12]. To mine human activities and routines from sociogeographic data in order to catch user's behavior, Cesario et al. [13] introduced a methodology for trajectory pattern mining and present a cloud-based framework for urban computing.

There are two main modeling approaches for trajectory data mining, which are grid-based model and graph-based model. The grid-based modeling approach is to first segment the map into grids and coordinate for these grids. With such operation, a grid-based model can be used to predict crowd density, taxi demand, traffic accident, and so on [14]. Jiang et al. [15] proposed a novel deep learning model called DeepCrowd by designing pyramid architectures and high-dimensional attention mechanism based on convolutional long short-term memory (LSTM) to address grid-based prediction for citywide crowd and traffic. Zhang et al. [16] explored zone clustering and how to utilize the interzone heterogeneity to improve the prediction, and they proposed a multilevel recurrent neural network (MLRNN) model to predict taxi demand. However, many traffic networks are graph-structured in nature, e.g., road network and subway network. The graph-based model introduces directed or undirected graph to utilize the topological structure of the road network for modeling [17]. Han et al. [18] proposed a digraph convolutional neural network to recognize the spatial traffic congestion by combining the graph Fourier transform with the directed graph decomposition method. They also proposed an attention mechanism-based digraph convolution network (ADGCN)-enabled framework to tackle the congestion recognition problem. They use a digraph-based convolution network to capture high-order spatial features [19]. These researchers made great achievements in many real-world applications, which gives us a lot of inspiration for our work.

B. Urban Anomaly Detection

Urban anomaly detection is an important part of smart cities and one of the guarantees for public safety. Traditional anomaly detection is performed manually, which is time-consuming, labor-intensive, and inefficient. Compared with traditional methods, data-driven methods have the advantages of real time and low cost, which have received wide attention for decades. Zhang et al. [20] made a comprehensive review of the state-of-the-art research on urban anomaly analytics and presented a comprehensive survey of issues on detecting and predicting techniques for urban anomalies. To accurately capture urban anomalies and timely alert for potential ones, Zhang et al. [21] proposed a two-step method to detect urban anomalies using multiple spatiotemporal data sources. Kong et al. [22] proposed a spatiotemporal cost combination-based framework for taxi driving fraud detection to analyze the causes of outlier trajectories combined with the perception of abnormal road environments. Zhang et al. [23] decomposed urban data into two parts and proposed a decomposition framework for detecting urban anomalies across spatiotemporal data. They employed the local outlier factor (LOF) method to detect anomalies by assigning an anomaly score to each data point. To collect the impacts of urban anomalies on multiple datasets, Liu et al. [24] proposed a novel end-to-end deep learning-based framework, namely, deep spatiotemporal multiple domain fusion networks. These works inspired us a lot in our framework. However, for the UOD problem, there are still several issues that need to be further explored, for example, how to define quantitative indicators of overtourism in different regions.

III. Overview

In this section, we introduce some definitions and concepts to express the framework more conveniently. Then, we provide a brief overview of the proposed framework.

A. Preliminary

Definition 1 (Grid Region): As a special detector of the traffic state section, the grid can collect traffic state information in a specific period [25]. We partition a city into an $m \times n$ grid map based on the longitude and latitude. Each grid is indexed by a unique ID.

Definition 2 (Time Interval): The smaller the time interval is, the easier it is to capture the change in a traffic state. However, if the time interval is too small, we cannot get enough trajectory information. We divide the day into 144 time intervals, each with an interval of 10 min, denoted as $t_1, t_2, \ldots, t_n$.

Definition 3 (Taxi Trajectory): A single taxi trajectory is represented as a sequence of GPS points with time intervals. It also contains the basic information about the taxi, such as plate number, speed, and passenger-carrying status.

Definition 4 (Traffic Congestion Index): The traffic congestion index indicates the congestion in a grid. Based on the road class in GIS data, we can get the speed limit of each road in the grid. We calculate the traffic congestion index of a particular road based on the taxi trajectory and the road speed limit. The traffic congestion index of the whole grid is the average of the traffic congestion index of all roads, which is calculated as follows:

$$C_{r,t_i} = \frac{1}{m}\sum_{j=1}^{m} c_{j,t_i} \tag{1}$$

where c_{j,t_i} represents the traffic congestion index of road j at time t_i.

Definition 5 (Number of Crowds): The number of crowds indicates the increase of taxi arrivals. The taxi trajectory data are processed into OD data, and when the passenger-carrying status of taxi changes from 1 to 0, it can be considered as a passenger drop-off indicating an increase in the number of crowds in the grid.

Definition 6 (GOM): The GOM is defined as a vector $\mathrm{GOM}_{t_i}(N, C)$, where N represents the number of crowds and C represents the traffic congestion index. We believe that urban overtourism can be considered only when traffic congestion is caused by the number of crowds [18]. From this perspective, we propose the GOM to detect overtourism on a grid level. Algorithm 1 demonstrates a more detailed calculation process.

Definition 7 (Overtourism Grid): An overtourism grid is a grid region that owns tourist attractions and has a GOM higher than the criteria for overtourism within the grid.

Problem Definition: Given the POI data and taxi trajectories data, we aim at finding overtourism grids at the next time interval.

B. Framework

We propose a novel UOD framework based on graph TCNs. The architecture of our proposed framework is shown in Fig. 3. First, we preprocess the data, including region division and time division. In addition, we screen out the regions containing tourist attractions according to the latitude and longitude of tourist attractions in POI data. Then, we propose the overtourism detection mechanism to screen the regions that overtourism at least once and construct the GOM graphs of the candidate regions. Next, we employ graph TCNs to deal with the spatiotemporal dependence and predict the future GOM graph at the next time interval. Finally, we calculate the urban overtourism scores based on the prediction results and get the final regions where overtourism occurs.

IV. Methodology

In this section, we first introduce the overtourism detection mechanism. Next, we introduce how to build the GOM graph. Then, we provide details for the graph TCNs, which are comprised of two stages of processing.

Algorithm 1 Calculate the GOM

Input: Taxi Trajectory Data D_r, Road Speed Limit Data V_{lim}, Time Intervals T, Road Network Data R in Grid G;
Output: GOMs in Grid G;

```
initialize: GOMs = ∅
for each time interval t_i in T do
    calculate the number of tourist arriving N_{t_i} through the Taxi Trajectory Data D_r
    for each road R_j in R do
        calculate the average road speed v
        v = (1/N) · Σ_{k=1}^{N} v̄_k
        calculate the average road congestion indexd c
        c = v_lim / v
    end
    calculate the grid congestion index C_{t_i}
    C_{t_i} = (1/M) · Σ_{l=1}^{M} c_l
    calculate the GOM in time interval t_i
    GOM_{t_i}(N, C)
    GOMs.append(GOM_{t_i})
end
return GOMs;
```

A. Overtourism Detection Mechanism

To identify whether overtourism occurs in a grid at a specific time interval, we propose the overtourism detection mechanism and give a quantitative indicator of overtourism. The specific steps are given as follows.

First, we calculate the average number of crowds and the average traffic congestion index across all time intervals in a grid, which can be combined into GOM. The sample mean of the GOM for the grid can be calculated as follows:

$$\overline{N} = \frac{1}{n}\cdot\sum_{i=1}^{n} N_{t_i}, \overline{C} = \frac{1}{n}\cdot\sum_{i=1}^{n} C_{t_i} \tag{2}$$

$$\mathrm{GOM}_{\mathrm{avg}} = \frac{1}{n}\cdot\left(\mathrm{GOM}_{t_1} + \cdots + \mathrm{GOM}_{t_n}\right) \tag{3}$$

where N_{t_i} and C_{t_i} represent the number of crowds and the traffic congestion index at t_i, respectively.

The Mahalanobis distance can be seen as a generalization of Euclidean distance [26], which can consider the relevance of the crowds and the traffic congestion index and is scale-invariant [27], [28]. Therefore, we use the Mahalanobis distance to detect GOMs in different time intervals, which is calculated as follows:

$$d_M\left(\mathrm{GOM}_{t_i}, \mathrm{GOM}_{\mathrm{avg}}\right) = \sqrt{D\cdot S^{-1}\cdot D^T} \tag{4}$$

$$D = \mathrm{GOM}_{t_i} - \mathrm{GOM}_{\mathrm{avg}} \tag{5}$$

where S is the covariance matrix of all GOMs in the same grid, which is calculated as follows:

$$S = \begin{bmatrix} \sigma\left(N_{t_i}, N_{t_i}\right) & \sigma\left(N_{t_i}, C_{t_i}\right) \\ \sigma\left(C_{t_i}, N_{t_i}\right) & \sigma\left(C_{t_i}, C_{t_i}\right) \end{bmatrix} \tag{6}$$

$$\sigma\left(N_{t_i}, N_{t_i}\right) = \frac{\sum_{i=1}^{n}\left(N_{t_i} - \overline{N}\right)^2}{n-1} \tag{7}$$

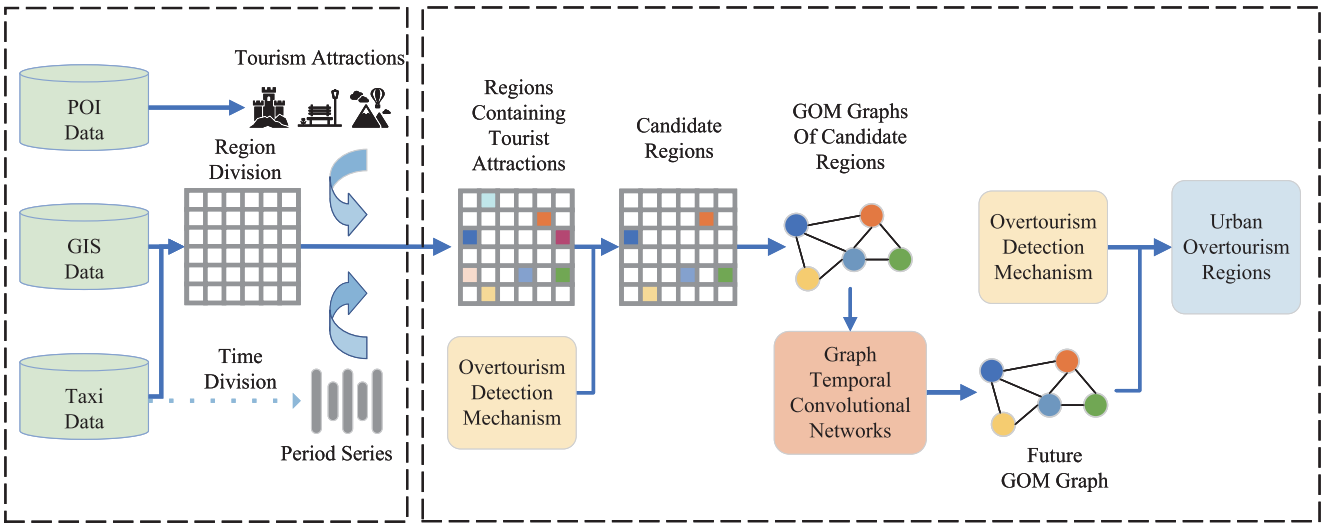


Fig. 3. Framework of UOD.

$$\sigma\left(N_{t_i}, C_{t_i}\right) = \frac{\sum_{i=1}^{n}\left(N_{t_i} - \overline{N}\right)\left(C_{t_i} - \overline{C}\right)}{n-1} \tag{8}$$

$$\sigma\left(C_{t_i}, N_{t_i}\right) = \frac{\sum_{i=1}^{n}\left(C_{t_i} - \overline{C}\right)\left(N_{t_i} - \overline{N}\right)}{n-1} \tag{9}$$

$$\sigma\left(C_{t_i}, C_{t_i}\right) = \frac{\sum_{i=1}^{n}\left(C_{t_i} - \overline{C}\right)^2}{n-1}. \tag{10}$$

Finally, we calculate the variations of GOMs in a series of consecutive time intervals. Specifically, we calculate the Mahalanobis distance between GOM_{t_i} and $\mathrm{GOM}_{\mathrm{avg}}$ over all time intervals, which is calculated as follows:

$$d_M\left(\mathrm{GOM}_{t_i}, \mathrm{GOM}_{\mathrm{avg}}\right) \geq \lambda \cdot \frac{\sum_{t_i=1}^{n} d_M\left(\mathrm{GOM}_{t_i}, \mathrm{GOM}_{\mathrm{avg}}\right)}{n}. \tag{11}$$

The left-hand side of the formula indicates the Mahalanobis distance between GOM_{t_i} and $\mathrm{GOM}_{\mathrm{avg}}$. The right-hand side indicates the average difference between every GOM over all time intervals and the average GOMs, which can represent the normal fluctuation. Equation (11) means the Mahalanobis distance between GOM_{t_i} and the normal fluctuation; λ represents the abnormal score and the higher value indicates the more abnormal.

In statistics, outliers refer to the values whose deviation from the mean is more than two standard deviations. Therefore, we choose λ to be 2, and $\lambda < 2$ represents normal fluctuation. In addition to the normal case, there are the following four abnormal cases when $\lambda > 2$.

1) When $N_{t_i} < \overline{N}$ and $C_{t_i} < \overline{C}$, the number of crowds and the traffic congestion index are both less than the average, which indicates that the area is unobstructed.
2) When $N_{t_i} < \overline{N}$ and $C_{t_i} > \overline{C}$, the traffic congestion index is greater than the average, but the number of crowds is less than the average, which indicates that there is heavy traffic but no overtourism.
3) When $N_{t_i} > \overline{N}$ and $C_{t_i} < \overline{C}$, the number of crowds is greater than the average, but the traffic congestion index is less than the average, which indicates that crowds are gathering in the area but no traffic congestion, the crowds' growth is still within the carrying capacity range, and there is no overtourism.
4) When $N_{t_i} > \overline{N}$ and $C_{t_i} > \overline{C}$, the number of crowds and the traffic congestion index are both greater than the average, the number of crowds increases substantially and leads to traffic congestion; thus, overtourism occurs.

Finally, we can screen the abnormal grids as the candidate regions when $N_{t_i} > \overline{N}, C_{t_i} > \overline{C}$. λ represents the urban overtourism score when $\lambda > 2$, and the higher value indicates the more serious overtourism. Algorithm 2 demonstrates a more detailed calculation process.

B. GOM Graph Generation

Compared with the whole urban regions, the regions where urban overtourism occurs only account for a small part. We first divide the city into grids and screen all the grids twice to find the grids where overtourism may occur. We screen the grid containing tourist attractions according to the latitude and longitude of tourist attractions in POI data. However, not all tourist attractions will be overtourism, such as some remote or unattractive tourist attractions that rarely become a destination for tourists. Therefore, in the second screening, we screen grids that have at least one overtourism according to the overtourism detection mechanism as the final research grids. Finally, we construct the GOM graph based on the screened girds, which allows us to focus on the grids where overtourism may occur and overcome the sparsity of urban overtourism.

According to Tobler's first law of geography, everything is related to everything else, but near things are more related than distant things [29]. Tourists tend to choose a nearby tourist attraction as their next destination after visiting the current tourist attraction. To extract the complex spatial relations, we consider building all candidate regions into a GOM graph

Algorithm 2 Overtourism Detection Mechanism

Input: All GOMs in Grid G;
Output: The urban overtourism GOMs set V in Grid G;

initialize: $V = \varnothing$
calculate the average GOM of a certain time intervals
$GOM_{avg} = (1/n) \cdot \left(GOM_{t_1} + \cdots + GOM_{t_n}\right)$
calculate the average of Mahalanobis Distance between every GOM and the average GOM
$\alpha = \left(\sum_{t_i=1}^{n} d_M\left(GOM_{t_i}, GOM_{avg}\right)\right)/n$
for *each* $GOM_{t_i}(N, C)$ *in GOMs* **do**
  calculate the Mahalanobis Distance between GOM_{t_i} and GOM_{avg}
  $\beta_{t_i} = d_M\left(GOM_{t_i}, GOM_{avg}\right)$
  if $\beta_{t_i} \geqslant 2 \cdot \alpha$ **then**
    if $N_{t_i} > \overline{N}, C_{t_i} > \overline{C}$ **then**
      V.append(GOM_{t_i})
    else
      continue;
    end
  else
    continue;
  end
end
return V;

$\mathcal{G} = (V, E)$, where V represents the set of nodes and E represents the set of edges. For each node $v_i \in V$, denoting the GOM, there are two associated values N_i and C_i representing the number of crowds and the traffic congestion index. For each edge $e_{ij} = (v_i, v_j)$, there is a weight a_{ij} representing the correlations between nodes. There are two steps to generate a GOM graph: select nodes to build the set V and then define the feature matrix X and the adjacency matrix A.

Specifically, we first determine the boundaries of urban regions and divide the regions into regular grids, as shown in Fig. 4(a). By analyzing GIS and POI data, we select the grids that contain tourist attractions, as shown in Fig. 4(b). Then, we screen M grids that overtourism at least once according to the overtourism detection mechanism as candidate regions and construct the set of nodes V, as shown in Fig. 4(c). Since M is much smaller than the whole number of urban grids, the grids where overtourism never occurs can be drastically reduced, and we can focus on those where overtourism occurs. Then, we use the GOM as node characteristic representation to construct the feature matrix X.

We use the reciprocal of geographical distance to represent connectivity between nodes and construct the adjacency matrix A. If two tourist attractions are far apart, tourists often do not go to another tourist attraction after visiting the current tourist attraction. In other words, there is less relation between the two nodes. The relationship between nodes can be defined by the following formula:

$$a_{ij} = \frac{1}{\text{distance}\left(v_i, v_j\right)}. \tag{12}$$

After the above steps, we can get a GOM graph $\mathcal{G}$ with its feature matrix X and adjacency matrix A, as shown in Fig. 4(d). The GOM of each grid will change over time, we use $\mathcal{G}_t$ to represent the GOM graph at t. Based on historical data, we can get a sequence of GOM graphs ordered by time $(\mathcal{G}_1, \ldots, \mathcal{G}_n)$.

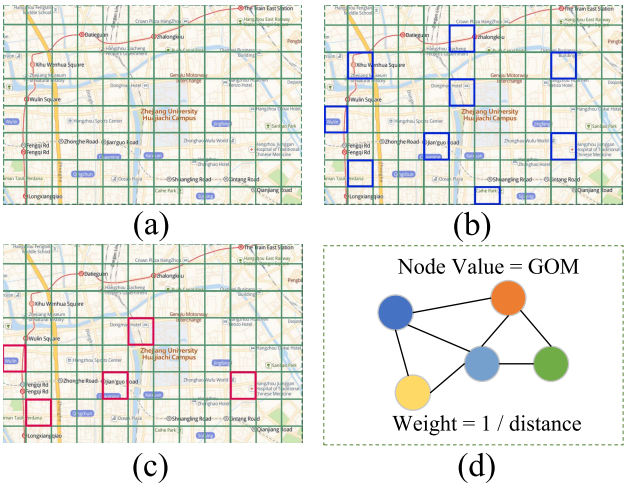


Fig. 4. Process of generating a GOM graph. (a) Boundaries of the studied urban areas are determined from GIS data and the urban region is divided into regular grids. (b) Screen the grids containing tourist attractions. (c) Screen the grids where overtourism may occur as candidate regions. (d) Build a GOM graph from the candidate regions and determine the weights.

C. Graph TCNs

To study the complex spatiotemporal factors of urban overtourism, we are supposed to mine historical data from the spatial and temporal dimensions respectively [30]. When looking into a single GOM graph $\mathcal{G}_t$, the weights on the edges represent the spatial relations between nodes. When considering a sequence of GOM graphs $(\mathcal{G}_1, \ldots, \mathcal{G}_n)$, there are temporal trends and periodicity that indicate a dynamic change in GOM graphs. We employ the graph TCNs to model the complex spatiotemporal relations of urban overtourism, as shown in Fig. 5.

Graph convolutional network (GCN) can generalize convolution operations from traditional Euclidean space data to graph-structured data and can consider the graph structural characteristics while considering the characteristics of nodes. Thus, we use GCN in [31] to extract the spatial relations in GOM graphs. Given an adjacency matrix A and the feature matrix X, the GCN constructs a filter on the GOM graph, which will map the nodes' relations into embedding space according to the values of nodes and the structure of the graph. To reduce the calculation complexity, the Chebyshev polynomial approximation [32] graph convolution can be rewritten as

$$g_\theta * \mathcal{G} = \sum_{k=0}^{k} \theta_k T_k\left(\frac{2L}{\lambda_{\max} - I_N}\right)\mathcal{G} \tag{13}$$

where the parameter θ is the vector of polynomial coefficients. $L = I_N - D^{-(1/2)} A D^{-(1/2)}$, where I_N is the identity matrix, D is the degree matrix, $D = \sum_j a_{ij}$, and $\lambda_{\max}$ is the largest eigenvalue of L. The recursive definition of the Chebyshev

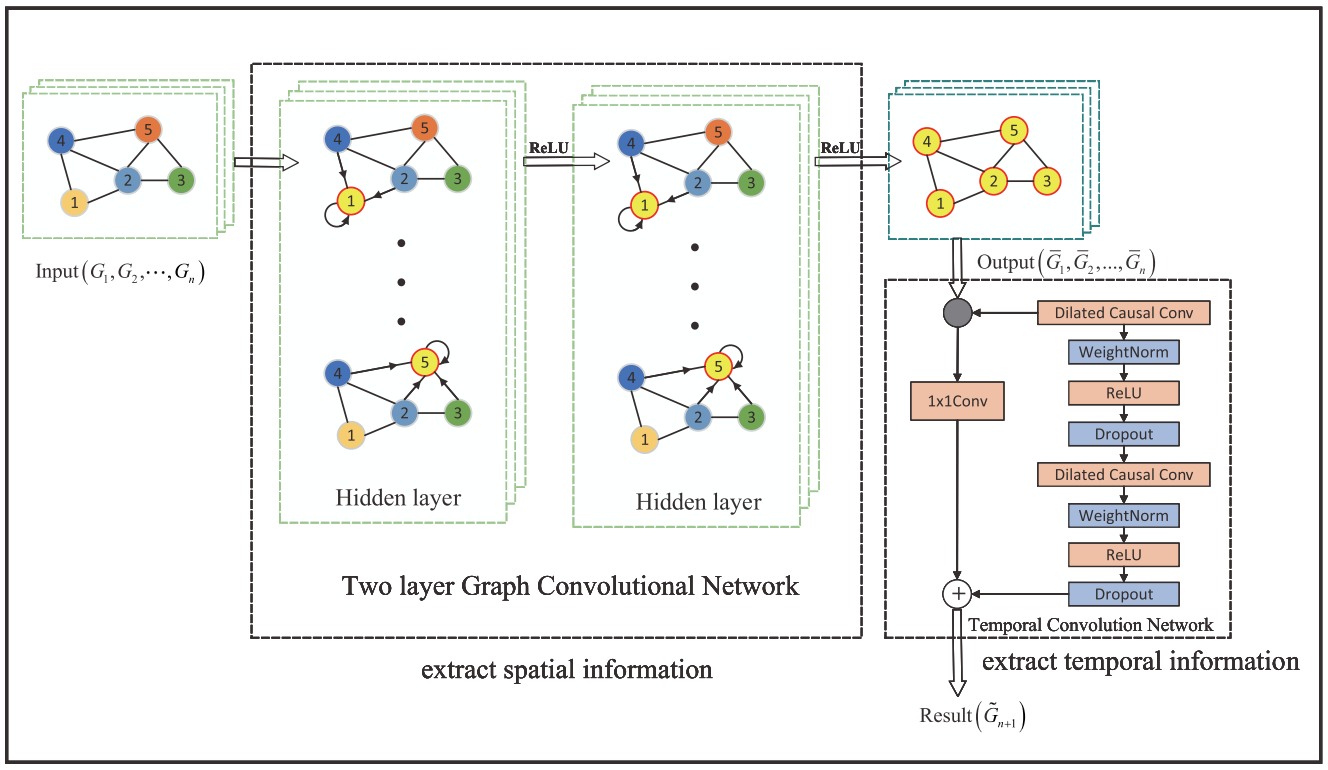


Fig. 5. Architecture of graph TCNs. The spatial relations are first extracted using a multilayer graph convolutional network to obtain the embedding of the nodes. To alleviate the overfitting, we added RelU layers as activation functions after each GCN layer. The output is a set of arbitrary graphs $(\overline{\mathcal{G}}_1, \ldots, \overline{\mathcal{G}}_n)$. The second part is a TCN, which extracts temporal correlations and predicts future GOM graph $\tilde{\mathcal{G}}_{n+1}$ at the next time interval.

polynomial is $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, where $T_0(x) = 1$ and $T_1(x) = x$. Therefore, the layerwise graph convolution process can be rewritten as

$$\mathcal{G}^{(l+1)} = \sigma\left(g\theta\mathcal{G}^l W^l + b^l\right) \tag{14}$$

where $\mathcal{G}^l$ is the output of the lth graph convolutional layer, W^l and b^l are the parameters of the lth graph convolution layer, and σ represents the activation function, here we use ReLU.

We use a two-layer GCN in our networks to map the nodes' features into embedding space. After each GCN layer, we add the dropout technique, which can reduce overfitting and achieve better embedding. Repeating this process for GOM graphs of all time intervals, we finally get a sequence of embedding ordered by time $(\overline{\mathcal{G}}_1, \ldots, \overline{\mathcal{G}}_n)$.

Time-series modeling usually adopts recurrent neural networks (RNNs) and related variants, such as LSTM and GRU. The architecture of the TCN is not only more accurate than these networks but also simpler and clearer. TCN can be processed in parallel on a large scale, which is faster in training and verification, and TCN avoids gradient vanishing and gradient exploding in RNN. Thus, we consider using TCN in [33] to further extract temporal relations between GOM graphs. TCN employs dilated convolutions to achieve an exponentially large receptive field. More formally, for an input sequence $(\overline{\mathcal{G}}_1, \ldots, \overline{\mathcal{G}}_n)$ and a filter $f : \{0, 1, \ldots, k-1\} \to \mathbb{R}$, the dilated convolution operation F on element s of the sequence is defined as

$$F(s) = \left(\overline{\mathcal{G}} *_d f\right)(s) = \sum_{i=0}^{k-1} f(i) \cdot \overline{\mathcal{G}}_{s-d\cdot i} \tag{15}$$

where d is the dilation factor, k is the filter size, and $s - d \cdot i$ accounts for the direction of the past. When $d = 1$, a dilated convolution reduces to a standard convolution.

The core idea of the residual block is to introduce a skip connection that directly skips one or more layers, which can effectively improve the performance of deep networks [34], as shown in Fig. 6. TCN employs a generic residual block, which contains a branch leading to a series of transformations $\mathcal{F}$. The outputs are added to the input of the block

$$o = \text{Activation}\left(\left(\overline{\mathcal{G}}_1, \ldots, \overline{\mathcal{G}}_n\right) + \mathcal{F}\left(\overline{\mathcal{G}}_1, \ldots, \overline{\mathcal{G}}_n\right)\right). \tag{16}$$

Furthermore, to account for discrepant input–output widths, TCN uses an additional 1×1 convolution to ensure that elementwise addition $\oplus$ receives tensors of the same shape.

The whole process can be represented by the following function:

$$\tilde{\mathcal{G}}_{n+1} = \mathcal{P}_m((\mathcal{G}_1, \ldots, \mathcal{G}_n), \Theta) \tag{17}$$

where Θ denotes the set of all trainable parameters in the GCN and TCN.

The loss function is defined as follows:

$$L(\mathcal{P}, \Theta) = \|\mathcal{P}((\mathcal{G}_1, \ldots, \mathcal{G}_n), \Theta) - \mathcal{G}_{n+1}\|^2. \tag{18}$$

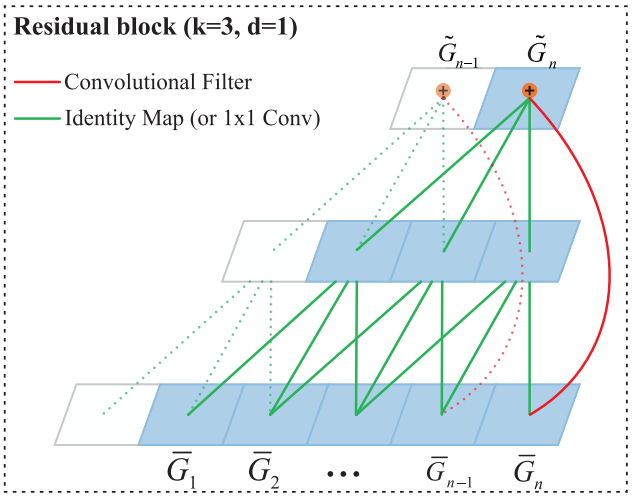


Fig. 6. Example of residual connection in a TCN. The green lines are filters in the residual function, and the red lines are identity mappings.

The goal of the training procedure is to minimize this loss function.

When the graph TCNs apply to forecast the GOM graph, there are the following advantages. First, the graph-structured representation can reduce redundant information and allow us to focus on regions where urban overtourism may occur. Second, GCN and TCN can fully mine historical data from spatial and temporal perspectives, respectively, to improve the accuracy of the results.

After we obtain the GOM graph at the next time interval, we can calculate the urban overtourism scores in each candidate region and obtain the urban overtourism regions at the next time interval.

V. Experiments

In this section, we conduct experiments on a real-world dataset containing multisource data to evaluate the proposed method. First, we introduce the dataset and data preprocessing. Then, we introduce the specific implementation and evaluation. To obtain a comprehensive assessment, we use five baselines and three indicators. We present experimental results and make a comprehensive analysis.

A. Data Preparation

The dataset contains taxi trajectory data collected by about 9000 taxis in Hangzhou within a month (September 18, 2017–October 17, 2017). The taxi trajectory data are uploaded by taxi every 40 s on average, and the average daily records from all taxis are more than 10 million pieces. The spatial and temporal distribution patterns of people and vehicles are embedded in the huge amount of taxi trajectory data. The area we discuss is a rectangle covering most of the main urban area of Hangzhou. In addition, we also obtained the POI dataset of Hangzhou through Gaode Map (https://ditu.amap.com/), which contains the longitude and latitude of tourist attractions.

We apply several preprocessing steps to adjust the real-world dataset to our UOD task. First, we divide the urban region into 41 $\times$ 45 regular grids, each with a size of about 2000 $\times$ 2000 m. Second, we screen the grids with tourist attractions according to the latitude and longitude of tourist attractions in POI data. Due to data security and privacy policies, it is difficult for us to get the exact number of crowds in each region. To some extent, the taxi trajectory data are a reasonable substitute. We can express the increase in the number of people by the behavior of passengers getting off the taxi in specific areas. When the passenger-carrying status of taxi changes from 1 to 0, it can be regarded as the arrival of the taxi passengers, which represents the increase in the number of crowds. In addition, the average vehicle speed can be calculated based on the taxi trajectory data, and the traffic congestion index of the road can be further calculated based on the speed limit of the road. The traffic congestion index of a grid is calculated by the average of all roads' traffic congestion index in the grid. The time interval t_i is set to 10 min because, if the time interval is too short, the grids cannot get enough taxi trajectory data to reflect the GOM of the grids. However, if the time interval is too large, the small changes in the overtourism cannot be captured.

B. Method Implementation

For specific experiments, we first describe how the overtourism detection mechanism detects overtourism in a grid. First, we calculate $\mathrm{GOM}_{\mathrm{avg}}$ for the grid at the weekend. Time intervals can be divided into three types, namely, weekdays, weekends, and holidays. On weekdays, most people need to work, and the value of GOM in tourist attractions is relatively low. During holiday peaks, popular tourist attractions are crowded, and the value of GOM in tourist attractions is relatively high. Therefore, it is not appropriate to calculate the average value of GOM on weekdays or holidays alone. Most residents choose to travel on weekends, and the number of crowds and traffic congestion index on general weekends can reflect the general city tourism situation; therefore, we use the average number of crowds and the average traffic congestion index on weekends as the standard situation to research urban overtourism.

Taking a grid as an example, we first calculate the average value of GOM in the grid, and the result is (15.36, 1.20). Fig. 7(a) and (c) shows the changes on weekdays and holidays, respectively, and the three lines indicate the number of crowds, traffic congestion index, and Mahalanobis distance. We normalize the values of crowds and traffic congestion index, which is calculated as (value min)/(max min), min and max are, respectively, minimum value and maximum value in an individual day. To see the change of the GOM more obviously, we represent the five cases mentioned in Section IV as -1 to 3, where -1 represents $N_{t_i} < \overline{N}$ and $C_{t_i} < \overline{C}$, 0 represents normal fluctuation, 1 represents $N_{t_i} < \overline{N}$ and $C_{t_i} > \overline{C}$, 2 represents $N_{t_i} > \overline{N}$ and $C_{t_i} < \overline{C}$, and 3 represents $N_{t_i} > \overline{N}$ and $C_{t_i} > \overline{C}$. The results are shown in Fig. 7(b) and (d).

In Table I, we list some typical abnormal GOM, including time interval t, $\mathrm{GOM}(N, C)$, the left part of (11) [i.e., $d_M(\mathrm{GOM}_{t_i}, \mathrm{GOM}_{\mathrm{avg}})$], the value of λ, the real case for abnormal GOM, and whether overtourism has occurred.

TABLE I
TYPICAL ABNORMAL GOMS IN A GRID

t	$GOM\ (N, C)$	left of Eq(11)	λ	GOM	Overtourism
4:00-4:10 on weekday	(2,1.02)	90.69	2.39	-1	No
7:00-7:10 on weekday	(1,1.63)	97.39	2.57	1	No
18:10-18:20 on weekday	(4,1.21)	77.09	2.04	1	No
20:00-20:10 on weekday	(29,1.42)	92.58	2.44	3	Yes
3:20-3:30 on holiday	(4,0.97)	77.13	2.04	-1	No
11:50-12:00 on holiday	(28,1.12)	85.75	2.26	2	No
13:30-13:40 on holiday	(29,1.29)	92.56	2.44	3	Yes
17:00-17:10 on holiday	(15,1.47)	85.82	2.27	3	Yes
20:30-20:40 on holiday	(29,1.59)	92.61	2.45	3	Yes

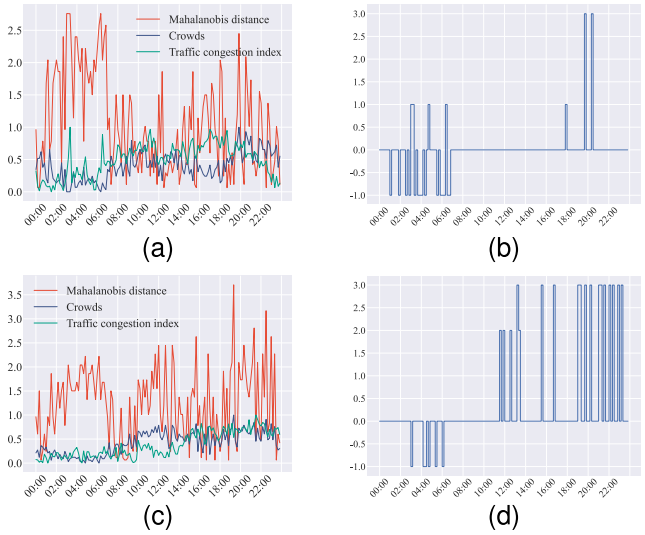


Fig. 7. Variations in crowds and traffic congestion index during weekdays and holidays. (a) Variations in crowds, traffic congestion index and Mahalanobis distance during weekdays. (b) Variations in the GOM cases during weekdays. (c) Variations in crowds, traffic congestion index and Mahalanobis distance during holidays. (d) Variations in the GOM cases during holidays.

We make a brief analysis of these four abnormal GOMs as follows.

1) From 4:00 to 4:10 on a weekday and from 3:20 to 3:30 on a holiday, people were mainly at rest, so the traffic congestion index and crowds were both very low.
2) From 7:00 to 7:10 and from 18:10 to 18:20 on a weekday, these time intervals were mainly morning and evening rush hours on weekdays, so there was traffic congestion.
3) From 11:50 to 12:00 on a holiday, despite the increase in the number of crowds, no traffic congestion occurred, indicating that the number of the crowds was still within the carrying capacity of the grid and did not overtourism.
4) During the rest of the time intervals, the number of crowds and the traffic congestion index were both more than the average, indicating overtourism.

Fig. 8 shows how the overtourism score changes in a grid on holidays.

C. Model Evaluation

We screen 42 grids containing tourist attractions and each grid has at least one overtourism according to the overtourism detection mechanism. Then, we construct the GOM graphs according to the proposed methods mentioned in Section IV. The value of each node is GOM. The weight between nodes is the reciprocal of the geographical distance. The valid duration of the entire dataset is 24 days. We recruit the first 19 days as the training set and the remaining five days are used for testing. We use the GOM graph of the past 60 min to predict the GOM graph of the next 10 min. To increase the number of samples, we use a 10-min moving window for data augmentation.

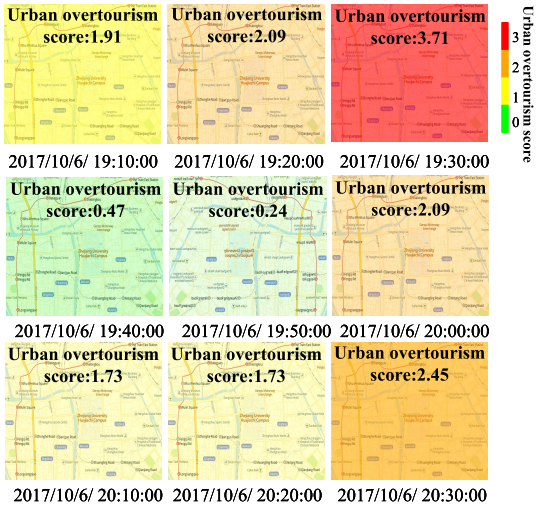


Fig. 8. Variations of urban overtourism score in a region during an hour and a half.

To verify the performance of the methods proposed in this article, we compare our methods with the following methods.

1) *Historical Average (HA):* HA takes the historical average as the prediction result.
2) *Autoregressive Integrated Moving Average Model (ARIMA) [35]:* ARIMA is a traditional, time-series forecasting model analyzed only at the statistical and theoretical levels.
3) *LSTM Neural Network [36]:* LSTM is a kind of temporal RNN specially designed to solve the long-term dependence problem of general RNNs.
4) *TCN [33]:* It is a convolution-based model for processing time series.
5) *ToGCN + Seq2Seq [14]:* Topological graph convolutional network avoids the vehicle to x (V2X) communication flash crowd situation, which is very similar to overtourism.

TABLE II
RESULTS OF METHODS COMPARED

Algorithms	Recall	RMSE	rRMSE
HA		3.8530	0.5134
ARIMA	0.380	3.8402	0.4761
LSTM	0.513	0.3354	0.1370
TCN	0.542	0.3228	0.1318
ToGCN+Seq2Seq	0.803	0.2264	0.0924
Ours	**0.841**	**0.2007**	**0.0820**

D. Trianing and Evaluation Metrics

The entire training process minimizes the loss function by adjusting the parameters of the model. The network parameters are trained from a random start, to optimize the objective function, we adopt the adaptive moment estimation (Adam) as the optimizer to perform all weights updated with a fixed learning rate and the learning rate is 0.001.

Evaluating UOD in the real-world settings is an open challenge since it is difficult to obtain a complete set of ground truths. Unlike traditional anomaly detection tasks, where the dataset contains labels such as anomalous events and anomalous states, urban overtourism is an anomaly that we define based on actual observations and statistical laws. Therefore, a direct comparison with other anomaly detection methods is not appropriate. In addition, some existing urban anomaly detection methods validate the accuracy of detection methods by prediction tasks [37]. Therefore, we calculate the root-mean-square error (RMSE) and relative RMSE (rRMSE) to evaluate the accuracy of the prediction results, which are defined as follows. Inspired by previous work [38], we also use the recall rate to evaluate the performance of our method. The recall rate can represent the ratio of detected abnormal events to true abnormal events, where TP indicates the number of detected abnormal events and FN indicates the model not detected in the abnormal events

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left\|\widetilde{\mathcal{G}}_{n+m} - \mathcal{G}_{n+m}\right\|^2} \tag{19}$$

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left\|\widetilde{\mathcal{G}}_{n+m} - \mathcal{G}_{n+m}\right\|^2}}{\frac{1}{M}\sum_{m=1}^{M}\mathcal{G}_{n+m}} \tag{20}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{21}$$

The results are shown in Table II. We can observe that the deep learning methods are significantly better than the statistical methods. Compared with the previous two statistical methods, LSTM and TCN are more advanced time-series models, which can better extract the temporal relations in historical data and their results have significantly improved. In fact, in addition to temporal relations, the overtourism situation in a tourist attraction is also influenced by other tourist attractions, i.e., there is a spatial relation between the occurrence of overtourism in tourist attractions. ToGCN + Seq2Seq considers both temporal and spatial relations and achieves better results. The method we proposed first uses GCN to model spatial relations and then uses TCN to model temporal relations, which can take full advantage of the spatiotemporal relations in historical data. From the results, we can observe that our methods consistently outperform the state-of-the-art baselines.

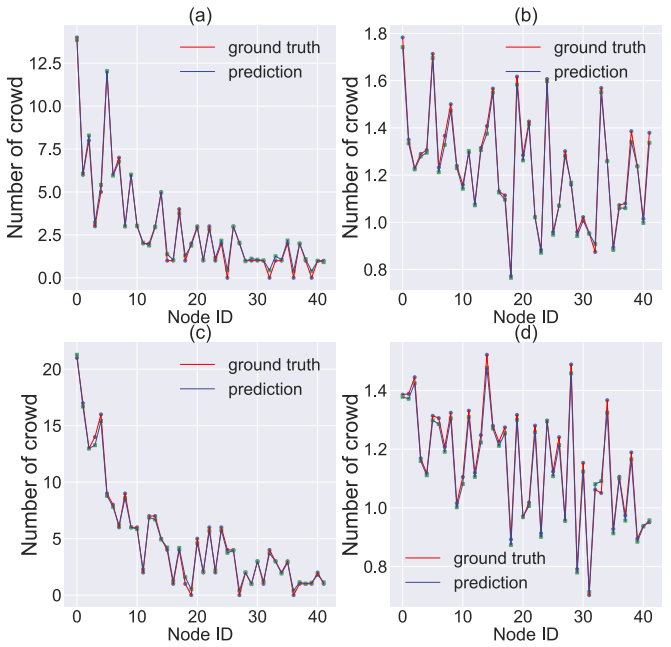


Fig. 9. Examples for comparing the ground-truth values with the predicted values in two randomly selected time intervals. (There are 42 nodes in total, and each node has two characteristics: crowd and traffic congestion index.) (a) and (b) Crowd and traffic congestion index at a certain time, respectively. (c) and (d) Crowd and traffic congestion index at another certain time, respectively.

We show the results of our methods with some examples in Fig. 9; specifically, we randomly choose two time intervals to display the prediction values and ground-truth values. According to the observed results, the prediction values and ground-truth values are close to each other, indicating the effectiveness of our method.

We evaluate the accuracy over different input sequence sizes. We use six time intervals to predict one time interval, which can be presented as $\{\{\mathcal{G}_{n+1}, \ldots, \mathcal{G}_{n+6}\} \rightarrow \widetilde{\mathcal{G}}_{n+7}\}$. We use the ground-truth situations and the prediction results for predicting more future time intervals, which can be presented as $\{\{\mathcal{G}_{n+1}, \ldots, \mathcal{G}_{n+6}, \widetilde{\mathcal{G}}_{n+7}\} \rightarrow \widetilde{\mathcal{G}}_{n+8}\}$. Although the prediction error is propagated to further predictions, the longer the prediction interval, the larger the error in the prediction result, the results are still acceptable, as shown in Table III.

GCN extracts the spatial information of the GOM graph by aggregating node features and edge features. To verify the effectiveness of embedding, we also do comparison experiments to investigate the effect of the number of GCN layers on the experimental results. We conducted experiments on no GCN, one-layer GCN, two-layer GCN, three-layer GCN, and four-layer GCN. The results of the comparison experiment are shown in Fig. 10(a). As the number of layers increases, the performance of the model will improve at first. We can observe that the result with a single-layer GCN is better than those without GCN, indicating that the graph convolutional network has a good ability to extract spatial dependency and the spatial information is meaningful to our task. However, the benefit of

TABLE III
EVALUATION ON PREDICTING MORE FUTURE TIME INTERVALS FOR OVERTOURISM

Prediction settings	Metrics	LSTM	TCN	ToGCN+Seq2Seq	**Ours**
$\{\mathcal{G}_{n+1},\ldots,\mathcal{G}_{n+6}\} \rightarrow \{\widetilde{\mathcal{G}}_{n+7}\}$	RMSE	0.3354	0.3228	0.2264	**0.2007**
	rRMSE	0.1391	0.1318	0.0924	**0.0820**
$\{\mathcal{G}_{n+1},\ldots,\mathcal{G}_{n+6},\widetilde{\mathcal{G}}_{n+7}\} \rightarrow \{\widetilde{\mathcal{G}}_{n+8}\}$	RMSE	0.3407	0.3293	0.2329	**0.2058**
	rRMSE	0.1418	0.1345	0.0951	**0.0850**
$\{\mathcal{G}_{n+1},\ldots,\mathcal{G}_{n+6},\widetilde{\mathcal{G}}_{n+7},\widetilde{\mathcal{G}}_{n+8}\} \rightarrow \{\widetilde{\mathcal{G}}_{n+9}\}$	RMSE	0.3472	0.3348	0.2387	**0.2137**
	rRMSE	0.1454	0.1367	0.0975	**0.0873**
$\{\mathcal{G}_{n+1},\ldots,\mathcal{G}_{n+6},\widetilde{\mathcal{G}}_{n+7},\widetilde{\mathcal{G}}_{n+8},\widetilde{\mathcal{G}}_{n+9}\} \rightarrow \{\widetilde{\mathcal{G}}_{n+10}\}$	RMSE	0.3515	0.3407	0.2420	**0.2223**
	rRMSE	0.1435	0.1391	0.0988	**0.0931**
$\{\mathcal{G}_{n+1},\ldots,\mathcal{G}_{n+6},\widetilde{\mathcal{G}}_{n+7},\widetilde{\mathcal{G}}_{n+8},\widetilde{\mathcal{G}}_{n+9},\widetilde{\mathcal{G}}_{n+10}\} \rightarrow \{\widetilde{\mathcal{G}}_{n+11}\}$	RMSE	0.3591	0.3492	0.2494	**0.2273**
	rRMSE	0.1467	0.1426	0.1018	**0.0952**

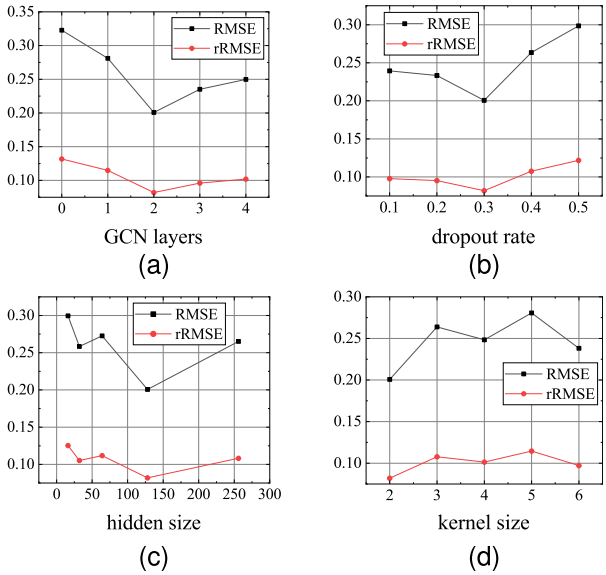


Fig. 10. Influence of some hyperparameters on the model. (a) Performance with respect to the size of GCN layer. (b) Performance with respect to the drop rate of TCN. (c) Performance with respect to the hidden size of GCN. (d) Performance with respect to the kernel size of TCN.

the number of GCN layers on the experimental results is not linearly increasing; when the number of GCN layers exceeds two, the performance of the model begins to degrade, probably because the embedding of nodes is very similar; and the deeper structure of GCN layers may cause oversmoothing when extracting spatial information [39]. In addition, we investigated the influence of some other hyperparameters on the model. Fig. 10(b) shows the impact of the dropout rate of TCN, a suitable dropout ratio greatly affects the performance of the model, and the best performance of the model is achieved when the dropout rate of TCN is 0.3. Fig. 10(c) shows the impact of the hidden size of GCN, and the hidden size of GCN is turned from [16, 32, 64, 128, 256]. The best performance of the model is achieved when the hidden size of GCN is 128. We can observe that the model performance shows a V-shaped trend as the hidden size increases. Fig. 10(d) shows the impact of the kernel size of TCN, and the best performance of the model is achieved when the kernel size of TCN is 2.

VI. FUTURE WORK

Our future work mainly includes the following aspects. The tourists' choices can be influenced by factors such as weather and unexpected events, and we will include this additional external information to further improve the accuracy of detecting overtourism. We will also try to use more novel and sophisticated deep learning models to mine hidden information from taxi trajectory data, such as the spatial dependence of traffic flow or spatiotemporal distribution of the crowd flow, to achieve more accurate and longer-term UOD. For example, the research on attention mechanisms has already shown a certain degree of success in data mining, and we consider introducing attention mechanisms into our future work. In addition, we will try to recommend tourist attractions and personalized travel routes for tourists by comprehensively considering the time cost, journey cost, personalized preferences, and urban overtourism. We hope to improve the quality of tourism for tourists while alleviating urban overtourism and promoting sustainable tourism development.

VII. CONCLUSION

In this article, we investigate the UOD problem. We propose a novel framework (UOD) based on graph TCNs for UOD. We define the UOD problem and propose the GOM to reflect the overtourism state of each grid. We first preprocess the data, including region division and time division. The proposed overtourism detection mechanism gives a quantitative definition of overtourism and we screen out the regions where overtourism may occur as candidate regions. Then, we construct the GOM graphs of the candidate regions. The graph TCNs can extract both temporal and spatial relations from historical data and predict the future GOM graph at the next time interval. Finally, we calculate the urban overtourism score in each candidate region according to the prediction results. Based on the multisource data in the real world, the experimental results show the effectiveness of the proposed methods.

REFERENCES

[1] M. Duignan, *Overtourism'? Understanding and Managing Urban Tourism Growth beyond Perceptions: Cambridge Case Study: Strategies and Tactics to Tackle Overtourism*. Madrid, Spain: United Nations World Tourism Organisation, Mar. 2019, vol. 2, pp. 34–39.

[2] M. Ntemi, I. Sarridis, and C. Kotropoulos, "An autoregressive graph convolutional long short-term memory hybrid neural network for accurate prediction of COVID-19 cases," *IEEE Trans. Computat. Social Syst.*, early access, Apr. 28, 2022, doi: 10.1109/TCSS.2022.3167856.

[3] H. Katsumi, W. Yamada, and K. Ochiai, "Generic POI recommendation," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., ACM Int. Symp. Wearable Comput.*, Sep. 2020, pp. 46–49.

[4] T. Murata and K. Totsuka, "Agent-based simulation for avoiding the congestions of tourists," in *Proc. 5th IEEE Int. Conf. Cybern. (CYBCONF)*, Jun. 2021, pp. 56–60.

[5] H. Mo, K. Yan, X. Zhao, Y. Zeng, X. Wang, and F.-Y. Wang, "Type-2 fuzzy comprehension evaluation for tourist attractive competency," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 1, pp. 96–102, Feb. 2019.

[6] X. Kong et al., "A federated learning-based license plate recognition scheme for 5G-enabled internet of vehicles," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8523–8530, Dec. 2021.

[7] C. Sun, S. Li, D. Cao, F.-Y. Wang, and A. Khajepour, "Tabular learning-based traffic event prediction for intelligent social transportation system," *IEEE Trans. Computat. Social Syst.*, early access, May 10, 2022, doi: 10.1109/TCSS.2022.3170934.

[8] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–41, May 2015, doi: 10.1145/2743025.

[9] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, Aug. 2022.

[10] X. Kong, K. Wang, M. Hou, F. Xia, G. Karmakar, and J. Li, "Exploring human mobility for multi-pattern passenger prediction: A graph learning framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 16148–16160, Sep. 2022.

[11] E. Cesario, C. Comito, and D. Talia, "An approach for the discovery and validation of urban mobility patterns," *Pervas. Mobile Comput.*, vol. 42, pp. 77–92, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S157411921630390X

[12] C. Comito, "NexT: A framework for next-place prediction on location based social networks," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106205. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070512030424X

[13] E. Cesario, C. Comito, and D. Talia, "Towards a cloud-based framework for urban computing, the trajectory analysis case," in *Proc. Int. Conf. Cloud Green Comput.*, Sep. 2013, pp. 16–23.

[14] H. Qiu, Q. Zheng, M. Msahli, G. Memmi, M. Qiu, and J. Lu, "Topological graph convolutional network-based urban traffic flow and density prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4560–4569, Jul. 2021.

[15] R. Jiang et al., "DeepCrowd: A deep model for large-scale citywide crowd density and flow prediction," *IEEE Trans. Knowl. Data Eng.*, early access, May 3, 2021, doi: 10.1109/TKDE.2021.3077056.

[16] C. Zhang, F. Zhu, Y. Lv, P. Ye, and F.-Y. Wang, "MLRNN: Taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8412–8422, Jul. 2022.

[17] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI*, vol. 33, 2019, pp. 922–929.

[18] X. Han, G. Shen, X. Yang, and X. Kong, "Congestion recognition for hybrid urban road systems via digraph convolutional network," *Transp. Res. C, Emerg. Technol.*, vol. 121, Dec. 2020, Art. no. 102877. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X20307774

[19] G. Shen, X. Han, K. Chin, and X. Kong, "An attention-based digraph convolution network enabled framework for congestion recognition in three-dimensional road networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 14413–14426, Sep. 2022.

[20] M. Zhang, T. Li, Y. Yu, Y. Li, P. Hui, and Y. Zheng, "Urban anomaly analytics: Description, detection, and prediction," *IEEE Trans. Big Data*, vol. 8, no. 3, pp. 809–826, Jun. 2022.

[21] H. Zhang, Y. Zheng, and Y. Yu, "Detecting urban anomalies using multiple spatio-temporal data sources," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 1–18, Mar. 2018, doi: 10.1145/3191786.

[22] X. Kong et al., "Spatial-temporal-cost combination based taxi driving fraud detection for collaborative internet of vehicles," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3426–3436, May 2022.

[23] M. Zhang, T. Li, H. Shi, Y. Li, and P. Hui, "A decomposition approach for urban anomaly detection across spatiotemporal data," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1–8.

[24] R. Liu, S. Zhao, B. Cheng, H. Yang, H. Tang, and T. Li, "Deep spatio-temporal multiple domain fusion network for urban anomalies detection," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, M. d'Aquin, S. Dietze, C. Hauff, E. Curry, and P. Cudré-Mauroux, Eds., Oct. 2020, pp. 905–914, doi: 10.1145/3340531.3411920.

[25] S. An, H. Yang, J. Wang, N. Cui, and J. Cui, "Mining urban recurrent congestion evolution patterns from GPS-equipped vehicle mobility data," *Inf. Sci.*, vol. 373, pp. 515–526, Dec. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025516304546

[26] P. C. Mahalanobis, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci.*, vol. 2, pp. 49–55, Jan. 1936.

[27] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proc. 21st Int. Conf. Adv. Geographic Inf. Syst.*, New York, NY, USA, Nov. 2013, pp. 344–353, doi: 10.1145/2525314.2525343.

[28] V. Torra and Y. Narukawa, "On a comparison between Mahalanobis distance and Choquet integral: The Choquet–Mahalanobis operator," *Inf. Sci.*, vol. 190, pp. 56–63, May 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025511006335

[29] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geography*, vol. 46, p. 234, Jun. 1970.

[30] Z. Huang, J. Zhang, L. Ma, and F. Mao, "GTCN: Dynamic network embedding based on graph temporal convolution neural network," in *Proc. 16th Int. Conf. Intell. Comput. Theories Appl. (ICIC)*, Bari, Italy. Berlin, Germany: Springer-Verlag, Oct. 2020, pp. 583–593, doi: 10.1007/978-3-030-60802-6_51.

[31] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.

[32] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 2, pp. 129–150, Mar. 2011.

[33] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[35] B. Williams and L. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov./Dec. 2003.

[36] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[37] L. Deng, D. Lian, Z. Huang, and E. Chen, "Graph convolutional adversarial networks for spatiotemporal anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2416–2428, Jun. 2022.

[38] M. Zhang, T. Li, H. Shi, Y. Li, and P. Hui, "A decomposition approach for urban anomaly detection across spatiotemporal data," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 6043–6049, doi: 10.24963/ijcai.2019/837.

[39] G. Li, M. Muller, A. Thabet, and B. Ghanem, "DeepGCNs: Can GCNs go as deep as CNNs?" in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9266–9275.

Xiangjie Kong (Senior Member, IEEE) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2004 and 2009, respectively.

He is currently a Professor with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou. Previously, he was an Associate Professor with the School of Software, Dalian University of Technology, Dalian, China. He has published over 160 scientific papers in international journals and conferences (with over 130 indexed by ISI SCIE). His research interests include urban computing, mobile computing, and computational social science.

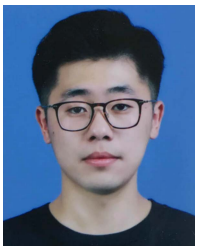
Dr. Kong is a Senior Member of China Computer Federation (CCF) and a member of Association for Computing Machinery (ACM).

Zhiqiang Huang received the B.Sc. degree in electronic and information engineering from Anhui Jianzhu University, Hefei, China, in 2020. He is currently pursuing the master's degree with the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China.

His research interests include urban computing and data analysis.

Guojiang Shen received the B.Sc. degree in control theory and control engineering and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively.

He is currently a Professor with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou. His current research interests include artificial intelligence, big data analytics, and intelligent transportation systems.



Hang Lin received the B.Sc. degree in computer science and technology from Wenzhou University, Wenzhou, China, in 2021. He is currently pursuing the master's degree with the School of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China.

His research interests include machine learning and data science.

Mingjie Lv received the Ph.D. degree in communication from the Communication University of China, Beijing, China, in 2017.

She is currently a Senior Research Officer with the Research Center for Intelligent Society and Governance, Zhejiang Lab, Hangzhou, China. She was the General Manager of the Strategic Cooperation Department, Wasu Media, Hangzhou, China. She has been the Deputy Chief Editor of "Exploring the way to a smart society: AI enabled social governance" and "Exploring the way to smart logistics: a landscape of technology-enabled logistics." She has published several articles in Guangming Daily and Chinese core journals.